SOR- and Jacobi-type Iterative Methods for Solving ℓ_1-ℓ_2 Problems by Way of Fenchel Duality[1]

Masao Fukushima[2]

July 17, 2010; revised February 4, 2011

Abstract

We present an SOR-type algorithm and a Jacobi-type algorithm that can effectively be applied to the ℓ_1-ℓ_2 problem by exploiting its special structure. The algorithms are globally convergent and can be implemented in a particularly simple manner. Relations with coordinate minimization methods are discussed.

Key words. ℓ_1-ℓ_2 problem, Fenchel dual, SOR method, Jacobi method.

1 Introduction

The purpose of this short article is to draw the reader's attention to the fact that the so-called ℓ_1-ℓ_2 problem can effectively be solved by iterative methods of SOR- or Jacobi-type by way of Fenchel duality. The ℓ_1-ℓ_2 problem is to find a vector $x \in \mathbb{R}^n$ that solves the following nonsmooth convex optimization problem:

$$\min_{x \in \mathbb{R}^n} \; f(Ax) + g(x), \tag{1}$$

where $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are defined by

$$\begin{aligned} f(s) &= \frac{1}{2}\|s - b\|_H^2, \\ g(x) &= \tau\|x\|_1, \end{aligned}$$

respectively, and $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\tau > 0$. Moreover, $\|\cdot\|_1$ denotes the ℓ_1 norm in $\mathbb{R}^n$ and $\|\cdot\|_H$ is the norm in $\mathbb{R}^m$ defined by $\|s\|_H = \sqrt{s^T H s}$ with a symmetric

[1]This work was supported in part by a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science.

[2]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail: fuku@i.kyoto-u.ac.jp

positive definite matrix $H \in \mathbb{R}^{m\times m}$. When $H = I$, the norm $\|\cdot\|_H$ reduces to the ℓ_2, or Euclidean, norm $\|\cdot\|_2$. This problem has recently drawn much attention in various application areas such as signal and image reconstruction and restoration [14].

The Fenchel dual [11] of problem (1) is stated as

$$\min_{y\in\mathbb{R}^m} \; f^*(-y) + g^*(A^T y), \tag{2}$$

where $f^* : \mathbb{R}^m \to \mathbb{R}$ and $g^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are the conjugate functions of f and g, respectively, and are given by

$$\begin{aligned} f^*(y) &= \frac{1}{2}\|y + Hb\|^2_{H^{-1}} - \frac{1}{2}\|Hb\|^2_{H^{-1}}, \\ g^*(t) &= \begin{cases} 0 & \text{if } \; t \in S := \{t \in \mathbb{R}^n \mid \|t\|_\infty \le \tau\} \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, ignoring the constant term, we may rewrite the dual problem (2) as follows:

$$\begin{array}{ll} \min & \frac{1}{2}\|y - Hb\|^2_{H^{-1}} \\ \text{s.t.} & -\tau e \;\le\; A^T y \;\le\; \tau e, \end{array} \tag{3}$$

where $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. In the following, we denote the columns of matrix A by $a^i \in \mathbb{R}^m$, $i = 1, \ldots, n$, which will be assumed to be nonzero throughout the paper. Then the constraints of problem (3) can be represented as

$$-\tau \;\le\; (a^i)^T y \;\le\; \tau, \quad i = 1, \ldots, n. \tag{4}$$

It may be worth mentioning that the above pair of dual problems can be derived in another way. First note that the KKT conditions for problem (3) can be written as

$$H^{-1}y - b + A\xi - A\eta = 0, \tag{5}$$

$$0 \le \xi \;\perp\; -A^T y + \tau e \ge 0, \tag{6}$$

$$0 \le \eta \;\perp\; A^T y + \tau e \ge 0, \tag{7}$$

where ξ and η denote the Lagrange multipliers associated with the right-hand and the left-hand inequality constraints in (3), respectively, and $a \perp b$ means vectors a and b are orthogonal. By (5), we have

$$y = -HA(\xi - \eta) + Hb. \tag{8}$$

This along with (6) and (7) yields

$$0 \leq \xi \perp A^T HA(\xi - \eta) - A^T Hb + \tau e \geq 0, \tag{9}$$
$$0 \leq \eta \perp -A^T HA(\xi - \eta) + A^T Hb + \tau e \geq 0. \tag{10}$$

It is then easy to observe that (9) and (10) comprise the KKT conditions for the following convex quadratic program:

$$\begin{array}{ll} \min & \frac{1}{2}(A(\xi - \eta) - b)^T H(A(\xi - \eta) - b) + \tau e^T(\xi + \eta) \\ \text{s.t.} & \xi \geq 0, \quad \eta \geq 0. \end{array} \tag{11}$$

It is not difficult to verify that any optimal solution of this problem satisfies

$$\xi_i \eta_i = 0, \quad i = 1, \ldots, n.$$

Therefore, by letting

$$x = \xi - \eta, \tag{12}$$

problem (11) can be rewritten as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}(Ax - b)^T H(Ax - b) + \tau e^T |x|,$$

where $|x| := (|x_1|, \ldots, |x_n|)^T$. This is precisely problem (1).

Note that the dual problem (3) has a unique optimal solution, whereas the primal problem (1) has an optimal solution but it is not necessarily unique. From (8) and (12), optimal solutions of problems (1) and (3) are related by

$$y = H(b - Ax). \tag{13}$$

A few words about notation: We let e^i denote the ith unit vector in $\mathbb{R}^n$, i.e., the ith column of the $n \times n$ identity matrix. The median of three real numbers α, β, γ is denoted by $\mathrm{mid}\{\alpha, \beta, \gamma\}$.

2 Algorithms

Hildreth's algorithm [6] and its Successive Over-Relaxation (SOR) modification [8] are classical iterative methods for solving strictly convex quadratic programming problems with inequality constraints. These methods use the rows of the constraint matrix just one at a time and act upon the problem data directly without modifying the original matrix in the course of the iterations; hence the name "row-action methods" [2]. This type of algorithms can be viewed as particular realizations of matrix splitting algorithms for linear complementarity problems, and their convergence properties have extensively been studied under general settings; see, e.g., [4, 7, 9, 10].

The quadratic program (3) has particular constraints that consist of pairs of linear inequalities, which we call interval constraints. The above-mentioned methods [6, 8] may naturally be applied to problems with interval constraints by treating each pair of inequalities as two separate inequalities. However this is by no means the best strategy. By exploiting the special feature of interval constraints, Herman and Lent [5] developed an extension of Hildreth's algorithm to deal with a pair of inequalities directly, see also [3]. Subsequently, an SOR version of the row-action method for interval constraints was presented in [12]. Moreover, a parallel Jacobi-type modification of the row-action method for interval constraints was proposed in [13].

In the following two subsections, we describe the SOR-type algorithm [12] and the Jacobi-type method [13], both of which fully exploit the special feature of the problem with interval constraints.

2.1 SOR-type algorithm

The SOR-type algorithm for solving (3) is stated as follows [12]:

Algorithm 1.
Initialization: Let $(y^{(0)}, x^{(0)}) := (Hb, 0) \in \mathbb{R}^m \times \mathbb{R}^n$ and choose a relaxation parameter $\omega \in (0, 2)$.
Iteration k: Choose an index $i_k \in \{1, \dots, n\}$ according to some rule and let

$$
\begin{aligned}
\alpha_{i_k} &:= (a^{i_k})^T H a^{i_k}, \\
\Delta^{(k)} &:= \frac{\tau - (a^{i_k})^T y^{(k)}}{\alpha_{i_k}}, \\
\Gamma^{(k)} &:= \frac{-\tau - (a^{i_k})^T y^{(k)}}{\alpha_{i_k}}, \\
c^{(k)} &:= \mathrm{mid}\left(x_{i_k}^{(k)}, \omega\Delta^{(k)}, \omega\Gamma^{(k)}\right), \\
x^{(k+1)} &:= x^{(k)} - c^{(k)} e^{i_k}, \\
y^{(k+1)} &:= y^{(k)} + c^{(k)} H a^{i_k}.
\end{aligned}
$$

Note that α_{i_k} are all positive, since H is positive definite and $a^{i_k} \neq 0$ by assumption.

Moreover, it is easily seen that $\Gamma^{(k)} < \Delta^{(k)}$ for all k.

The algorithm generates two sequences $\{y^{(k)}\} \subseteq \mathbb{R}^m$ and $\{x^{(k)}\} \subseteq \mathbb{R}^n$. It can easily be shown that they are related by the formula

$$y^{(k)} = H(b - Ax^{(k)}), \quad k = 0, 1, 2, \ldots \tag{14}$$

Under the assumption that the selection of indices $\{i_k\}$ follows the almost cyclic rule, i.e., there exists an integer $N > 0$ such that $\{1, 2, \ldots, n\} \subseteq \{i_k, i_{k+1}, \ldots, i_{i+N}\}$ for all k, it is shown that the whole sequence $\{y^{(k)}\}$ converges to the unique solution y^* of problem (3) [12, Theorem 4.3] and the rate of convergence is N-step linear in the sense that, for some constant $\rho \in (0, 1)$, the inequality

$$\|y^{(k+N)} - y^*\|_{H^{-1}} \;\leq\; \rho \, \|y^{(k)} - y^*\|_{H^{-1}}$$

holds for all k large enough [12, Theorem 4.4]. In view of the relations (13) and (14), we can deduce that any accumulation point of the sequence $\{x^{(k)}\}$ is an optimal solution of problem (1).

2.2 Jacobi-type algorithm

The Jacobi-type algorithm for solving (3) is stated as follows [13]:

Algorithm 2.
Initialization: Let $(y^{(0)}, x^{(0)}) := (Hb, 0) \in \mathbb{R}^m \times \mathbb{R}^n$ and choose a relaxation parameter $\omega > 0$.
Iteration k: (i) For $i = 1, \ldots, n$, let

$$\begin{aligned}
\alpha_i &:= (a^i)^T H a^i, \\
\Delta_i^{(k)} &:= \frac{\tau - (a^i)^T y^{(k)}}{\alpha_i}, \\
\Gamma_i^{(k)} &:= \frac{-\tau - (a^i)^T y^{(k)}}{\alpha_i}, \\
c_i^{(k)} &:= \operatorname{mid}\left(x_i^{(k)}, \omega\Delta_i^{(k)}, \omega\Gamma_i^{(k)}\right).
\end{aligned}$$

(ii) Let

$$\begin{aligned}
x_i^{(k+1)} &:= x_i^{(k)} - c_i^{(k)}, \quad i = 1, \ldots, n, \\
y^{(k+1)} &:= y^{(k)} + H \sum_{i=1}^{n} c_i^{(k)} a^i.
\end{aligned}$$

Note that we always have $\alpha_i > 0$ and $\Gamma_i^{(k)} < \Delta_i^{(k)}$, $i = 1, \ldots, n$. Moreover, step (i) can be implemented in parallel for $i = 1, \ldots, n$. Like Algorithm 1, this algorithm generates two sequences $\{y^{(k)}\} \subseteq \mathbb{R}^m$ and $\{x^{(k)}\} \subseteq \mathbb{R}^n$ that satisfy

$$y^{(k)} = H(b - Ax^{(k)}), \quad k = 0, 1, 2, \ldots \tag{15}$$

Let

$$\begin{aligned} \alpha_{ij} &:= (a^i)^T H a^j, \quad i, j = 1, \ldots, n \ (i \neq j), \\ \theta_i &:= \frac{2}{\alpha_i} \sum_{j=1, j\neq i}^{n} |\alpha_{ij}|, \quad i = 1, \ldots, n, \end{aligned}$$

and define

$$\begin{aligned} \bar{\omega}_i &:= \min\left\{\frac{1}{\theta_i}, \frac{3}{2+\theta_i}\right\}, \quad i = 1, \ldots, n, \\ \bar{\omega} &:= \min_{1\leq i\leq n} \bar{\omega}_i. \end{aligned}$$

It is shown [13, Theorem 3.8] that if the relaxation parameter ω is chosen to satisfy the condition $\omega \in (0, \bar{\omega})$, then the sequence $\{y^{(k)}\}$ generated by the Jacobi-type algorithm converges to the unique solution y^* of problem (3). Moreover, from the relations (13) and (15), any accumulation point of the sequence $\{x^{(k)}\}$ is an optimal solution of problem (1).

3 Discussion

It is well-known that the so-called row-action methods are 'dual' to the coordinate minimization methods, which search for a next iterate along some coordinate selected possibly in an almost cyclic manner. In fact, for a given $x^{(k)}$, the exact minimizer of the objective function $f(Ax) + g(x)$ of problem (1) along the ith coordinate can explicitly be computed as $x^{(k)} - c^{(k)} e^i$, where

$$c^{(k)} = \operatorname{mid}\left(x_i^{(k)}, \frac{\tau - (a^i)^T H(b - Ax^{(k)})}{(a^i)^T H a^i}, \frac{-\tau - (a^i)^T H(b - Ax^{(k)})}{(a^i)^T H a^i}\right).$$

In view of the relation (14), we find that the (exact) coordinate minimization method for problem (1) is equivalent to Algorithm 1 with relaxation parameter $\omega = 1$.

Coordinate minimization methods for nonsmooth optimization problems have been studied by a number of authors. Auslender [1, Chapter VI, Section 1] established convergence of the method with relaxation parameter $\omega \in (0,2)$ by assuming the objective function is strongly convex. Note, however, that the last assumption is not satisfied by problem (1).

Tseng [15] studied a (block) coordinate descent method for solving a nonsmooth optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} \; f(x) + \sum_{j=1}^{J} f_j(x^j),$$

where $x^j \in \mathbb{R}^{n_j}$ are sub-vectors that compose the vector $x \in \mathbb{R}^n$, i.e., $n = n_1 + \cdots + n_J$, f is a smooth function and f_j are nonsmooth convex functions. Tseng and Yun [17] considered the problem

$$\min_{x \in \mathbb{R}^n} \; f(x) + P(x),$$

where f is a smooth function and P is a nonsmooth convex function, and propose (block) coordinate gradient descent methods with some stepsize rule based on a descent condition. The above two problems contain problem (1) as a special case. In [15, 17], without assuming the function f to be convex, it is shown that the (block) coordinate gradient descent methods generate a sequence $\{x^{(k)}\}$ whose accumulation point is a stationary point of the corresponding minimization problem. For related results, see [16, 18].

4 Conclusion

We have presented an SOR-type algorithm and a Jacobi-type algorithm for solving the ℓ_1-ℓ_2 problem. These algorithms exploit the special structure of the problem and can be implemented in a very simple manner. The algorithms generate two sequences; one is convergent to the unique optimal solution of the dual problem, while any accumulation point of the other sequence is an optimal solution of the original ℓ_1-ℓ_2 problem.

Although no numerical results are given here, the behavior of the algorithms can be estimated from the extensive computational experience reported in [12, 13] for convex quadratic programs with interval constraints, since the algorithms in [12, 13] are essentially the same as those considered in this paper. In particular, the numerical results in [12, 13] show that the algorithms typically exhibit a linear rate of convergence.

References

[1] A. Auslender, *Optimisation Méthodes Numériques*, Masson, Paris, 1976.

[2] Y. Center, Row-action methods for huge and sparse systems and their applications, *SIAM Review* **23** (1981), pp. 444–466.

[3] Y. Censor and A. Lent, An iterative row-action method for interval convex programming, *Journal of Optimization Theory and Applications* **34** (1981), pp. 321–353.

[4] R.W. Cottle, J.-S. Pang and R.E. Stone, *The Linear Complementarity Problem*, Academic Press, San Diego, CA, 1992.

[5] G.T. Herman and A. Lent, A family of iterative quadratic optimization algorithms for pairs of inequalities, with application in diagnostic radiology, *Mathematical Programming Study* **9** (1979), pp. 15–29.

[6] C. Hildreth, A quadratic programming procedure, *Naval Research Logistics Quarterly* **4** (1957), pp. 79–85.

[7] A.N. Iusem, On the convergence of iterative methods for symmetric linear complementarity problems, *Mathematical Programming* **59** (1993), pp. 33–48.

[8] A. Lent and Y. Censor, Extensions of Hildreth's row action method for quadratic programming, *SIAM Journal on Control and Optimization* **18** (1980), pp. 444–454.

[9] Z.-Q. Luo and P. Tseng, Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem, *SIAM Journal on Optimization* **2** (1992), pp. 43–54.

[10] O.L. Mangasarian, Solution of symmetric linear complementarity problems by iterative methods, *Journal of Optimization Theory and Applications* **22** (1977), pp. 465–485.

[11] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[12] Y. Shimazu, M. Fukushima and T. Ibaraki, A successive over-relaxation method for quadratic programming problems with interval constraints, *Journal of Operations Research Society of Japan* **36** (1993), pp 73–89.

[13] T. Sugimoto, M. Fukushima and T. Ibaraki, A parallel relaxation method for quadratic programming problems with interval constraints, *Journal of Computational and Applied Mathematics* **60** (1995), pp. 219–236.

[14] J.A. Tropp and S.J. Wright, Computational methods for sparse solution of linear inverse problems, *Proceedings of the IEEE, Special Issue on Applications of Sparse Representation and Compressive Sensing* **98** (2010), pp. 948–958.

[15] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, *Journal of Optimization Theory and Applications* **109** (1991), pp. 475–494.

[16] P. Tseng, Approximation accuracy, gradient methods, and error bound for structured convex optimization, *Mathematical Programming* **125** (2010), pp. 263–295.

[17] P. Tseng and S. Yun, A coordinate gradient descent method for nonsmooth separable minimization, *Mathematical Programming* **117** (2009), pp. 387–423.

[18] S. Yun and K.-C. Toh, A coordinate gradient descent method for ℓ_1-regularized convex minimization, Singapore-MIT Alliance, Singapore, April 2008 (revised January 2009), *Computational Optimization and Applications*, to appear.